DOCUMENT RESUME

ED 138 205 HE 008 870

TITLE	Student Evaluation of Teaching. Criteria for the Evaluation, Support, and Recognition of College Teachers Number 2.
INSTITUTION	Michigan Univ., Ann Arbor. Center for Research on Learning and Teaching.
SPONS AGENCY	Fund for the Improvement of Postsecondary Education (DHEW), Washington, D.C.
PUB DATE	Jun 76
NOTE	6p.; Special publication of the Fund Associates in National Project III.
AVAILABLE FROM	National Project III, Center for Research on Learning and Teaching, University of Michigan, 109 E. Madison, Ann Arbor, Michigan 48109
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	Case Studies; *College Faculty; *College Instruction; Evaluation Criteria; *Evaluation Methods; *Higher Education; Information Sources; Measurement Instruments; Policy; Problems; Reliability; *Student Evaluation of Teacher Performance; Validity
IDENTIFIERS	National Project III (FIPSE)

ABSTRACT

Teaching is an omnibus profession, but each teacher is a self-sufficient individual and many yardsticks are needed to measure competence in this role. The evaluation of a teacher should, in principle, be bound to what students learn and to the attitudes and values they hold over the long haul. In practice, however, we tend to separate teaching from learning. Each of the participating schools in the National Project III uses some form of teacher evaluation by students, and from these collective experiences it is possible to be explicit about the issues and problems generated by such arrangements. These matters range from broad policy questions to technical decisions pertaining to the evaluating instrument and the handling of data. The report is a brief summary of these issues, and most of the examples are taken from Fund Associates schools. Sources for more detailed information are listed. (Author/MSE)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

CRITERIA FOR THE EVALUATION, SUPPORT, AND RECOGNITION OF COLLEGE TEACHERS

A SPECIAL PUBLICATION OF THE FUND ASSOCIATES IN NATIONAL PROJECT III, PREPARED AT THE CENTER FOR RESEARCH ON LEARNING AND TEACHING, THE UNIVERSITY OF MICHIGAN

Number II June, 1976

ED138205

Student Evaluation of Teaching

Teaching is an omnibus profession, but each teacher is a self-sufficient individual and many yardsticks are needed to measure competence in this role. The evaluation of a teacher should, in principle, be bound to what students learn and to the attitudes and values they hold over the long haul. In practice, however, we tend to separate teaching from learning—"I did a good job of teaching today; but whether my students learned anything was up to them." This is not a common response but it does illustrate the natural interest of teachers toward being evaluated in terms of **what they do as teachers**.

Each of the participating schools in National Project III uses some form of teacher evaluation by students and from our collective experience we can be quite explicit about the issues and problems generated by such arrangements. These matters range from broad policy questions to technical decisions pertaining to the evaluating instrument and the handling of data. The present report is a brief summary of these issues and most of our examples are taken from the Fund Associates. Interested readers are encouraged to write to the Fund Associates in National Project III (see list at end of text) for procedural specifics. The projects at Purdue and Kansas State University are especially worthy of attention.

The Institutional Context

The responsibility of the home institution is to evaluate fairly the individual members of the faculty, and this is a far more complicated task than the development and use of student ratings of teachers. Nevertheless, at many schools these procedures are not independent events and this section will identify some of the larger issues that tie-in student ratings to institutional policies.

Recognition and promotion on the basis of merit is often strongly defended, but this principle probably lacks force in many postsecondary institutions. It is not particularly difficult to find some good things to say about most teachers but unless the dimensions of merit have been explicitly set forth, we may have nothing more than window dressing for a seniority system of advancement and recognition. Promotion to a tenured rank involves a prediction of the career contribution of the teacher to the aims and goals of the institution. What are these institutional values to which the aspiring young teacher must conform?

Institutional Differences

ISSUE 1: To what extent should the evaluation of teachers reflect the priority values of the home institution?

Outstanding teachers share common characteristics of excellence, regardless of the type of school. Even so, there are real differences in the pressures experienced by teachers at different institutions. The statewide SUNY* system is examining its procedure for granting special awards for outstanding teaching and will determine how this mode of recognition is perceived by the faculties on the different campuses. A close look at a similar arrangement (special awards to teachers) is underway at The University of Michigan.* The SUNY-Oswego* project is a good example of combining the values and preferences of the individual teacher with the standards set by the department chairperson.

Course-specific Differences

ISSUE 2: How might an evaluation procedure balance institutional needs with the distinctive factors in the teaching task of the individual teacher?

We tend to talk about teaching as a general skill, but fair and valid evaluation requires special attention to the specific conditions of subject matter, teacher, student characteristics and special conditions affecting the environment for learning. These influences derive from different combinations of factors from one teacher to the next or from one course to another. Most teachers accept the accountability principle insofar as they have confidence in the criteria and the measures used for evaluating their performance. Their sensitivity to these matters is quite legitimate.

Care must be taken to establish the pertinent criteria for each instructional setting and to judge the teacher within this context. The clarity and relevance of the teacher's course objectives, for example, should carry considerable weight in teacher evaluation, as should the ability to organize course content into a productive hierarchy and to assess student performance in a manner that supports rather than hinders learning. A good teacher must be able to provide instructional materials relevant to the objectives of the course, to tutor, to counsel, to excite students and, finally, to serve as an exemplar or model for the attitudes and values germane to a particular area of research, teaching and public service. These are some of the dimensions of good teaching but each is manifest in a distinctive way by the idiosyncratic teacher/course combination.

*See Criteria I for summaries of the different Fund Associate programs referred to in the present report. Request copies from: Center for Research on Learning and Teaching, University of Michigan, 109 E. Madison, Ann Arbor, MI 48109.

Teaching and/or Research and Service?

ISSUE 3: How much weight is assigned to the evaluation of teaching in concert with the other contributions of a faculty member?

There is a difference between the rather specific responsibilities of a classroom teacher and his or her broader functions as a member of the faculty. Institutional recognition often derives from the more visible activities of committee work, administrative responsibilities, scientific and scholarly engagement, publications, community services, leadership in professional organizations, and the many other activities that gain attention and favorable reaction from the larger community. The effect of these activities may or may not contribute to the quality of instruction received by students in the classroom. What is best for the institution or the teacher's professional development is, in the long run, usually best for the students, but in the meantime, certain aspects of classroom teaching are important for the here-and-now student.

Good teaching is not necessarily correlated, plus or minus, with conformity to administrative criteria. Deans, teachers, and students each view the educational scene from their own vantage point and a fair system for faculty evaluation would be to openly examine these criteria. A distinctive feature of the project at the University of Illinois* is to elicit opinions of students toward their educational program—their "major." How, for example, do majors in chemistry regard the undergraduate program they experience? Faculty judgments are also obtained in this rather intensive analysis of the quality of a particular sequence or pattern of courses. Information from each of the several sources is weighed by a specially appointed task force responsible for the review and the development of recommendations.

The Dimensions for Evaluation

An experienced teacher will usually find that the average overall rating received from students does not change dramatically from term to term. Greater attention, therefore, is given by the teacher to those questionnaire items which provide specific diagnostic information about particular features of a course. These data are probably more useful as a means for helping a teacher improve a course than as a source of evidence for purposes of merit recognition and promotion.

Evaluating Content and/or Method

ISSUE 4: By what means might the rating scale separate **course** evaluation from the **personal style** of the teacher?

• A distinction must be made between evaluating the teacher as a person and the course as an organized program of study. These are not, of course, independent factors since a dull teacher can destroy an otherwise exciting body of knowledge and a charismatic teacher can breathe life into dreary textbook knowledge. A rating scale must do more than scale a teacher's "popularity" since these happiness scores may be quite unrelated to the educational impact of a course. Nevertheless, it is not at all uncommon for a student to like "the course" better than "the teacher" and these discriminations should not be obscured by the rating instrument.

Judgments by Peers and Supervisors

ISSUE 5: Are the factors best evaluated by peers and supervisors clearly distinguished from dimensions best judged by students?

The specific task of teaching is only one component in the full inventory of faculty responsibilities. Attention to housekeeping chores, for example, might be important to administrators and to colleagues, but whether or not the classroom teacher performs these logistical duties neatly and on time is of little immediate consequence to students. A teacher's reputation takes shape among his or her peers from the accumulation of incidents and comments during the normal course of departmental and institutional affairs. One's classroom style may not be known or given much weight by fellow faculty members who sense there is no single model for good teaching.

The criteria for successful teaching are not posted on a bulletin board or encoded in a set of bylaws. These standards grow and take form as traditions of the department develop and accommodate to the necessary variations in teaching style. Academia treasures individuality, but it takes courage, even so, to march to a different beat than the one given by the dean, the department chairperson, or the power structure within the department. After National Project III has moved further along, we will prepare a report to analyze the evaluation of teaching by one's self, by peers, and by those who administer an educational program.

Student Ratings of Teachers

The remainder of Criteria II will deal almost exclusively with student ratings of teachers. Teachers usually want to know how their students evaluate the main features of a course and the way it was taught. If these ratings are obtained within a climate of cooperation and mutual respect, they are a valuable source of information about the quality of instruction. We will outline the considerable research and development activity in the area of student ratings and will indicate at least some of the problems and issues. References to specific studies will be omitted since these can be found in the publications cited in the Bibliography.

Student Purposes in Evaluation

ISSUE 6: Do students make accurate observations of those features of a course that are significant to them?

The freedom for students to elect different courses means little if choice is based on trivial information. As "consumers" they want to know a great deal more about a course than the content area it covers. Questionnaire data can indicate what students judge to be important: instructional objectives, flexibility to pursue particular topics, how class time is spent, the supporting resources for instruction, evaluation procedures, grading standards, frequency and nature of tests, and finally, those idiosyncratic characteristics of the course

and the teacher that might make some difference to some students.

Mandatory or Voluntary Use of Rating Instruments

ISSUE 7: What schedule of student ratings (mandatory or voluntary; every class or some classes) gives optimum results?

The matter of "overexposure" to a specific rating system is important. Basically, this comes down to the question of mandatory versus voluntary use of the evaluation procedure on the part of the teacher. For an evaluation system to remain effective, students must be willing to give carefully considered opinions. We might expect the average quality of responses to decline considerably if all classrooms, every term, are saturated with the requirement to complete the rating forms. As a compromise, various intermittent schedules could be established but, in any case, decisions as to the frequency of questionnaire distribution should serve the best interests of the teachers.

ISSUE 8: What is the primary purpose of the rating instrument: evaluation and/or diagnosis?

The specific items used in the rating form must be consistent with the purpose for which the results are intended—information for course selection by students or diagnostic analysis for the teacher. The data that go forward through administrative channels for merit review may require yet a further set of questionnaire items. This selective use of items is an extremely important matter. When a teacher seeks diagnostic information, he or she might select questions by which students could point out the weaker aspects of a course or teaching method. If these results are then used for merit review, the evaluative system is working at cross purposes. If given a choice of questions for an administrative assessment, a teacher will tend to emphasize known strong points and perhaps inadvertently gloss over inadequacies that might damage his or her teaching image. If the teacher feels student ratings do not give an accurate reflection of his or her instructional plan and performance, it would hardly seem appropriate to forward these findings as evidence of professional competence. An inflexible or highly prescribed evaluation system can penalize the inventive or unconventional teacher. Such systems tend to converge teaching styles: to reward conformity to a preestablished template as to what is good teaching.

A major development in the current technology of student ratings is to include only a few compulsory items for evaluating the global or general characteristics of the teacher and the course. The following "core" items are included in the CAFETERIA system under current development and use at Purdue University:*

This instructor motivates me to do my best work.
Course assignments are interesting and stimulating.
This instructor explains difficult material clearly.
Overall, this course is among the best I have ever taken.
Overall, this instructor is among the best teachers I have known.

The bulk of each CAFETERIA instrument, however, consists of items selected by the individual teacher from a "catalog" of 200 or more items. The instructor can select up to 40 items (some of which can be self-constructed) referring to particular aspects of the course for which feedback is desired. This capability not only adapts the instrument to a variety of courses and teaching styles but it involves the teacher in a process he or she can shape or influence. Flexibility appears to be a major factor in gaining faculty acceptance and adoptions of CAFETERIA services.

The IDEA (Instructional Development and Effectiveness Assessment) system at Kansas State University* allows the teachers to identify a unique profile of objectives from a list of 10 different statements. Students rate their progress toward these aims in comparison to other classes. They also evaluate the instructor (20 items), the course demands (4 items), and complete five "self-rating" items plus eight demographic-type questions. The instructor receives a detailed report giving the frequency distribution of responses for all items and can differentiate the findings in terms of the best match between the objectives of the course, the size of the class, and particular instructional approaches employed. One of the more distinctive features of the KSU arrangement is its carefully worked out system of reference points; sets of norms which allow the teacher to take into account, for example, five different levels of student motivation and four different class sizes.

Developing a Useful Yardstick

ISSUE 9: What is the role of the measurement/evaluation specialist in the development of local evaluation instruments?

"Judge not, lest ye be judged." This admonition has now gone full circle and teachers, who have been passing judgement on student performance for countless years, are now being evaluated by students. Unfortunately, the quality of the measuring instruments is uneven. Examination and testing bureaus help faculty to develop discriminating procedures for evaluating students, but teachers are frequently "graded" by unreliable homemade instruments. Considering the complexity and the not-too-subtle threats of student rating systems, it is mandatory for the institution to develop procedures that meet at least minimum standards of consistency, accuracy, and fairness to individual teachers.

It is easier to measure height and weight than to assess intelligence or subject-matter knowledge. It is even more difficult to assess the **interaction** between teacher and students. The nature of the task requires that each rating form be capable of reflecting functionally relevant characteristics of a given teacher in a given course at a given institution, with the foreknowledge that the perception of these conditions will differ widely among students. The teacher needs no technical consultant to know how to ask students if they enjoyed the course. If, however, the questionnaire becomes rather complex and if certain quantitative treatments are to be applied to students' responses, e.g., norms, percentile ranks, etc., the teacher, the department, and the college are advised to obtain some guidance from persons knowledgeable about the several alternatives and pitfalls of such procedures.

The Main Factors to Which Students Respond

ISSUE 10: Does the choice of times cover the main factors in instruction?

Over the years literally thousands of different items have been included in teacher rating forms. By means of a rather sophisticated statistical treatment—factor analysis—it is possible to determine which of these items seem to cluster together, i.e., which ones tend to measure a common dimension or feature of instruction. The four principal factors seem to be:

(1) Skill. This is the most powerful general factor since **over half** the typical rating form items have substantial loadings in this factor while less than a tenth of the items can be placed in any one of the other clusters or categories. As far as the instructor is concerned, the skill factor is the most important dimension to be assessed by students. A sample question would be: The instructor gives clear explanations.

(2) Rapport, e.g., The instructor treats students with respect.

(3) Organization, e.g., The instructor uses class time well.

(4) Overload/Difficulty, e.g., The instructor has made the course sufficiently difficult to be stimulating.

As mentioned on page 1, the true measure of the teacher is the impact on students. A good rating scale should, therefore, include items which enable the students to indicate, in various ways, the impact value of the course. Given a free choice, the teacher may select more items aimed at the "teaching" than the "learning" (impact) side of things.

Norms

ISSUE 11: Are the available norms applicable and fair to each of the different teacher-course combinations?

On the face of it, the interpretation of student ratings would seem to be more meaningful if students' responses could be compared to established norms. There are, nevertheless, some problems in this arrangement. A normative comparison must be compatible with the situation-specific characteristics of a given teacher and a given course. If the use of a student rating scale is voluntary on the part of the teachers, it is questionable that "institutional" norms should be developed from a self-selected sample of the faculty.

The research findings show that student ratings show a "halo effect," that is, more often than not, students seem to like their teachers and this "bias" shows up when their ratings are averaged. If these ratings are then statistically transformed into a "normalized" frequency distribution, the teacher receives a somewhat distorted score since half of the teachers who contribute to these norms will be placed "below average." The straightforward use of norms at KSU is simply to present the frequency distribution and the teacher can then make a direct comparison of his or her ratings with those of other members of the faculty who teach comparable courses.

The value of student ratings is increased if the instructor will focus attention on specific items and on patterns revealed across item responses rather than trying to derive a gross "teaching index" score. If the established norms are limited to a total score, they may have the effect of pressing individual members of the faculty to teach in ways that are calculated to yield a "high grade." This is directly comparable to the competitive misdirections so frequently seen when students work for grades rather than to acquire and to understand a body of knowledge. If a teacher wants to know how he or she stands overall, simply ask two questions of students:

1. How do you rate this course, overall, in comparison with other courses you have taken?
2. How do you rate this teacher in comparison with other teachers you have had?

Follow-up

ISSUE 12: Can teachers pull themselves up by their own bootstraps?

Where can a teacher—as a teacher—go for help? We have never heard of a pedagogical crisis center for college professors and most of us make quite a point of hiding the troubles we have with our classes. Good teaching is taken for granted and most institutions simply have not found it necessary to establish counseling mechanisms to assist the troubled teacher—other than the department chairperson, one's spouse, or Kelsey's Bar. However, it is perfectly sensible to seek information as to how best to interpret rating-scale response. This does not mean that the teacher is "in trouble."

One of the better examples of a follow-up service is at Kansas State University where a knowledgeable second person helps to guard teachers against drawing false conclusions from rating data, to resolve conflicts and to choose among alternative types of corrective action. The KSU follow-up arrangement has become the main gateway to the larger program of faculty development.

Validity

ISSUE 13: Does feedback from students bring about significant changes in the classroom performance of a teacher?

The research evidence is, again, inconclusive. A teacher who wants to know the reaction of students to various features of a course will certainly be sensitive to the information received; it has salience and immediate validity. Upon receiving a set of completed ratings from a course, teachers will frequently tally, examine, fume, and purr. Even without reference to external norms, it may be apparent to the teacher that he or she is still not skilled, for example, in the management of group discussion. On the other hand, the teacher may be pleased to find that certain new features of the course were well received by students. This is valid information.

The one-shot set of questionnaire returns is less valuable to the teacher or to the administration than the accumulated ratings over time. This more stable average is a better indicator of a teacher's characteristic strengths and weaknesses but whether the teacher can do much about correcting deficiencies is quite another question. Knowing one is overweight does not lead automatically to weight loss; the quiet science professor is unlikely to become a charismatic spellbinder simply because this style seems preferred by students. After five or more years of teaching, it is difficult to change one's habits of speaking, to become more or less the authoritarian teacher, to be more receptive to contrary student opinions, to relax one's standards for grading, and so on. Nevertheless, most teachers will at least try to reduce the dissonance between their teaching habits and how these seem to be perceived by students.

The present statement notes current developments toward better procedures for evaluating teachers, especially as judged by students. However, no paper-and-pencil instrument yet devised can do complete justice as an evaluating procedure for the college teacher. Knowing the strengths and the limitations of these formal arrangements is one important step for guarding against their misuse.

Recent Comprehensive Reports

Specific articles and research reports are available in most of the current professional journals in higher education and educational research.

Bergquist, W. H., & Phillips, S. R. **A handbook for faculty development.** Washington, DC: The Council for the Advancement of Small Colleges, 1975.

Centra, J. A. **Strategies for improving college teaching.** Prepared by ERIC Clearinghouse. Published by American Association for Higher Education, Washington, DC, December 1972.

Doyle, K. O., Jr. **Student evaluation of instruction.** Lexington, MA: Lexington Books, 1975.

Eble, K. E. **The recognition and evaluation of teaching.** Washington, DC: Published by the American Association of University Professors, 1971.

Gage, N. L. "Student ratings of college teaching: Their justification and proper use." In Allen, Melnik, and Peele (Eds.), **Reform renewal reward.** Amherst, MA: Clinic to Improve University Teaching, University of Massachusetts, 1975. Pp. 121-135.

Kulik, J. A., & McKeachie, W. J. "The evaluation of teachers in higher education." In F. N. Kerlinger (Ed.), **Review of Research in Education, Vol. 3.** Itaska, IL: F. E. Peacock Publishers, Inc., 1975.

Miller, R. I. **Developing programs for faculty evaluation.** San Francisco: Jossey-Bass Publishers, 1974.

Fund Associates
National Project III

Ball State University, 2000 University Avenue, Muncie, IN 47306
Project Director: Jerry J. Nisbet, Coordinator of University Evaluations

Bucknell University, Lewisburg, PA 17837
Project Director: Wendell I. Smith, Provost

Burlington County College, Pemberton-Brown Mills Road, Pemberton, NJ 08068
Project Director: Lee L. Schroeder, Director of Measurement and Evaluation

Earlham College, National Road West, Richmond, IN 47374
Project Director: Gerald R. Bakker, Department of Chemistry

University of Illinois, 205 South Goodwin, Urbana, IL 61801
Project Director: Charles J. McIntyre, Director, Office of Instructional Resources

Kansas State University, Fairchild Hall, Manhattan, KS 66506
Project Director: Richard E. Owens, Office of Educational Improvement

The University of Michigan, 109 E. Madison, Ann Arbor, MI 48109
Project Director: Stanford C. Ericksen, Center for Research on Learning and Teaching

Ohio Wesleyan University, Delaware, OH 43015
Project Director: Stephen C. Scholl, Director, Instructional Development

Purdue University, ENAD, Room 402, West Lafayette, IN 47907
Project Director: Warren Seibert, Director, Office of Instructional Services, Measurement and Research Center

State University of New York, Central Administration, 99 Washington Avenue, Albany, NY 12210
Project Director: Charles B. Neff, Assistant Chancellor for Special Projects

State University of New York, Oswego, Oswego, NY 13126
Project Director: David J. King, Department of Psychology

Activity which is the subject of this report was supported in whole or in part by the Fund for the Improvement of Postsecondary Education, Department of Health, Education and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the Fund and no official endorsement by that agency should be inferred.

I prepared the original draft which was then reviewed by the Fund Associates. This final statement could not, however, reflect all of the comments and criticisms and I must, personally, assume responsibility for the omissions and the views herein presented. SCE